

DOCUMENT RESUME

ED 229 394

TM 830 284

AUTHOR Conrad, Kendon J.
TITLE A Demonstration of Causal Modeling in the Utilization of Program Implementation Measures.
SPONS AGENCY National Inst. of Mental Health (DHHS), Rockville, Md.
PUB DATE Apr 83
GRANT PHS-MH-15589-04
NOTE 24p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, April 11-15, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Analysis of Covariance; *Compensatory Education; Data Analysis; Data Collection; Educational Objectives; *Evaluation Methods; Evaluation Utilization; Measurement Techniques; Parent Participation; Preschool Education; *Program Evaluation; *Program Implementation; Young Children
IDENTIFIERS *Causal Inferences; *Causal Models; Chicago Child Child Parent Center IL; Classroom Observation Rating Scale; Comprehensive Tests of Basic Skills

ABSTRACT

This study evaluated the implementation of and the outcomes from a local child parent center compensatory education program. It intended to contribute to a better understanding of one phase of compensatory education; and, in the process, it proposed to develop a relatively simple and practical evaluation strategy which would verify the implementation of a program and relate the implementation measures to the outcome measures. The evaluation strategy included practical methods for modeling the program, gathering data, and analyzing data. Causal modeling techniques were used to relate implementation measures to outcome measures. The strategy was then analyzed regarding its usefulness as an evaluation design which would measure program implementation and provide explanatory power. (Author/PN)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED229394

X This document has been reproduced as
received from the person or organization
originating it.

Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

A Demonstration of Causal Modeling in the Utilization
of Program Implementation Measures*

Kendon J. Conrad
University of Illinois at Chicago

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

K. J. Conrad

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

Presented at the Annual Meeting of the American Educational
Research Association, Montreal, April, 1983

* Although the author assumes sole responsibility for its present form
and content, this paper has benefitted from the comments of Robert F.
Boruch and Maurice J. Eash. The author is also indebted to Leo Levy and
NIEH Grant #PHS MH 15589-04 for support during the completion of this
work.

Abstract

This study evaluated the implementation of and the outcomes from a local child parent center compensatory education program. The evaluation strategy included practical methods for modeling the program, gathering data, and analyzing data. Causal modeling techniques were used to relate implementation measures to outcome measures. The strategy was then analyzed regarding its usefulness as an evaluation design which would measure program implementation and provide explanatory power.

In the typical quasi-experimental evaluation, two non-equivalent groups, experimental vs. control, are compared post treatment with attempts made to ensure pre-treatment equivalence via analysis of covariance (ANCOVA). Lord has noted that, "No logical or statistical procedure can be counted on to make proper allowances for uncontrolled pre-existing differences between groups" (p. 305). Additionally, ANCOVA has been found, typically, to underadjust for pre-treatment differences (Campbell & Boruch, 1975); thereby not simply providing equivocal results, but equivocal results which indicate that treatment was harmful.

Campbell and Boruch suggest that this situation be remedied by the greater use of experimental designs (random assignment). Additionally, Magidson and Sobom (1982) have demonstrated that structural equation modeling techniques allow more precise estimates of pre-treatment differences between groups, i.e., the true score differences between groups. This technique requires multiple measures of independent and dependent variables. The problems with these two remedies for quasi-experimental designs are that (1) usually it is difficult to implement and maintain random assignment, and (2) that local evaluators often lack the wherewithal to obtain multiple measures and perform the intricate statistical analyses required by structural equation models.

Are there any more alternatives? Recent research suggests that the emphasis on outcomes, while especially important in traditional psychological research, is somewhat off-target in evaluation studies (Charters & Jones, 1973; Erikson, 1979; Fullan & Pomfret, 1977; Leinhardt, 1980). The study at hand finds that an emphasis on measuring program processes (implementation) sheds light on the question of

program value by first asking not "Did the experiment work?" but "Was there an experiment?" In other words, programs which are found to successfully implement the intended treatment are inherently more valuable and, other things being equal, have a theoretically better chance of causing the intended outcomes.

The Need for Process Measures

In spite of the growing awareness of the need to study processes in order to account for outcomes, actual evaluations of school programs have typically continued to concentrate on outcomes (usually achievement alone). Boruch's recent recommendations to Congress advocating the routine measurement of implementation (Boruch, Cordray, Pion, & Leviton, 1981; Boruch & Cordray, 1980) were founded on recent research such as that conducted by Forgione, Kaplan, and Orland (Note 1) in "Evaluation of Compensatory Education Programs: Problems, Promising Strategies and Recent Trends." The authors studied the types of evaluation strategies employed in evaluating Title I programs by 32 school districts in eight states. They found that evaluation primarily meant achievement testing to both states and local districts and that evaluation was viewed most frequently as a mechanical exercise unrelated to other administrative activities. It was further found by David (Note 2) in her study of evaluation utilization at the local level that local evaluations were not used either as a basis for judging the effectiveness of Title I programs or as guides for program decision-making.

With the development of improved methods of measuring program implementation (Leithwood & Montgomery, 1982; Rubin, Stuck, & Revicki, 1982; Gersten, Carnine, & Williams, 1982; DeVault, Harnishfeger, & Wiley, 1978) there arises the need for practical methods of utilizing

those implementation measures in local program evaluations. The present study demonstrates one such method using basic causal modeling techniques to relate measures of program implementation to measures of program outcome.

General Objectives

The general purpose of this study was to verify the implementation of the intended treatments in the Chicago Child Parent Center (CPC) compensatory education program (Conrad & Eash, 1982). With this general purpose, the study intended to contribute to a better understanding of one phase of compensatory education; and, in the process, it proposed to develop a relatively simple and practical evaluation strategy which would verify the implementation of a program and relate the implementation measures to the outcome measures.

It should be noted that the evaluation strategy proposed here did not propose to offer solutions to the much discussed problem of measurement error (e.g., Boruch & Gomez, 1979). Rather, it simply attempted to apply recent advances in evaluation technology employing basic causal modeling techniques to a local program evaluation.

Compensatory Education Programs

Compensatory education programs have, of course, been studied before; most notably, the Head Start program. The negative findings from the evaluation of Head Start began a continuing discussion of methodology in evaluating large social programs (Westinghouse/Ohio University, 1969). Many problems flawed the Head Start evaluation, but the main problems of concern here are that Head Start treatment was poorly specified and that implementation of treatment was poorly verified. The evaluation strategy described here addressed these two

concerns by delineating the steps needed to specify and verify the implementation of treatment.

Methods

With this brief background, this paper turns to a description and demonstration of an evaluation strategy which employed implementation measures and related them to outcome measures with the purpose of aiding causal inference. The data used in the demonstration of the model were taken from the Chicago Child Parent Center Study (Conrad & Eash, 1982).

A causal program model was constructed. The model was based the intentions of the program administrators as to the processes and outcomes which made up the program. The use of a model greatly facilitated our ability to think about and analyze program implementation and outcomes.

Specifically, the CPC model for improving academic achievement assumed that criterion performance of students on achievement tests was affected by the following classroom variables:

- A reduced class size, which was intended to allow increased individualization and increased interaction with the teacher and teacher-aide, thus producing a more child-centered classroom
- A highly structured basic skills program with increased student evaluation
- An abundant supply of learning materials resulting in an enriched educational environment
- Parent involvement in the school program

Figure 1 depicts a theoretical model of the CPC program.

Insert Figure 1 about here.

Appropriate instruments to measure processes and outcomes were selected. In the CPC study, the "Classroom Observation Rating Scale" (Walberg and Thomas, 1974) was adapted to assess classroom treatment. It was believed that the CORS was appropriate for this evaluation because the themes covered by the CORS were similar to the intended processes of the CPC program. The CORS was developed following the compensatory education movement and emphasized themes which were similar to those intended by the philosophy of compensatory education. Regarding the reliability of the CORS, Walberg and Thomas said:

Statistical analysis of the resulting data showed that the total score on the scales is highly reliable with internal consistencies above .95 whether obtained through questionnaire or observation. Moreover, they discriminate between open and traditional classes in American and British schools in advantaged and disadvantaged areas at high levels of statistical significance (1974).

Three scales were delineated a priori from the CORS: 1) child-centeredness; 2) presence of evaluation of student achievement; 3) enriched environment; and a fourth scale, parent involvement, was constructed specifically for this evaluation (see Table 1).

Insert Table 1 about here.

The outcome measure used was the Prereading Achievement Scale of the Comprehensive Tests of Basic Skills, Level A, Form S (CTBS, 1973). CTBS reports internal consistency reliabilities of .96 for the Total Prereading Scale.

A cadre of trained observers was obtained. In the CPC study a cadre of school and community representatives from participating public schools and faculty and graduate assistants from the university were trained to collect data. A single, one hour training session was held in which items were read and examples of situations which would apply to particular items were presented. The faculty and graduate assistants who participated in the classroom observations had previously worked on similar projects and were considered experienced and expert observers.

Measures of classroom processes were obtained from a random sample of 13 treatment and 6 control classes. The sample population of the CPC implementation study was composed of classrooms of students from age cycle 5 in four CPC's and affiliated public schools as well as from two schools not in any way affiliated with CPC's. All schools were located in the same school district and served similar disadvantaged populations of whom the overwhelming majority was black.

The major problem with this study was the small sample of classrooms and the consequent low power to detect effects. The reader should bear this in mind when effect sizes are reported as non-significant at $p < .05$. On the other hand, the fact that class means were used on the dependent measure made this estimate relatively stable.

Class means on measures of student outcomes were obtained from the sample classrooms. It was necessary to obtain class means because the classroom was the appropriate unit of analysis. In this

analysis, the 5-year-old students were sorted by room, and their Prereading Achievement scores were averaged.

Results

Outcomes were related to treatment to determine the strength of the relationship. This step alone is the crucial analysis of the typical outcomes evaluation.

Insert Figure 2 about here.

Figure 2 reveals that when Prereading Achievement was regressed on CPC Program (dummy coded: CPC=1, non-CPC=0) the path coefficient was non-significant. In the typical outcomes evaluation this would indicate that the program had no effect. However, the following path models demonstrated that a great deal of information about how the CPC Program affected classroom processes and about how classroom processes affected Prereading achievement was absent from this analysis.

The magnitude of the process measures was observed to determine whether they were being implemented at intended levels. In the CPC implementation study, it was decided that a mean item response above 2.5 on a 4 point scale would be considered strong evidence and below 2.5 weak evidence of the presence of the various classroom processes. 2.5 was chosen simply because it was the mid-point on the 4 point scale.

Insert Table 2 about here.

Table 2 reveals that the CPC classrooms scored above 2.5 on the

Evaluation and Enriched Environment scales, indicating strong implementation of these classroom variables. However, Child-centeredness was found to be below 2.5 but was significantly higher ($p < .05$) than non-CPC classrooms. In this case, the criterion of statistical significance was thought more compelling and valid than the arbitrary criterion of the 2.5 mid-point, and this finding was interpreted as being significant. Thus, although both CPC and non-CPC classrooms were found to be low in Child-centeredness, CPC classrooms were interpreted as being significantly higher in this attribute. Parent Involvement was below 2.5 and showed no statistically significant difference between CPC and non-CPC classrooms.

Processes were related to treatment to determine if the intended program processes were actually being implemented to a significantly greater degree than in conventional programs. This was the "program testing stage" of the analysis. In other words, this analysis tested whether the experimental program was a significant improvement over the conventional program on measures of the specified variables. In the CPC study the process measure (CORS Total) was regressed on treatment (dummy coded 1=CPC, 0=non-CPC). The regression estimated the strength of the relationship between the process measures and being a CPC classroom. The left side of Figure 3 depicts the strength of the overall regression of Treatment on Process Total, the sum of the four scales of the CORS. Figure 4 depicts the results of the regressions of separate process scales on Treatment. This model involved four separate equations, the regressions of each scale of the CORS on treatment. The path coefficients on the left in Figure 4 reveal that CPC treatment caused statistically significant differences in "child-centeredness" and

"evaluation" but not in "enriched environment" and "parent involvement."

Insert Figures 3 and 4 about here.

Outcomes were related to processes to determine the strength of the relationship. This was the "theory testing stage" of the analysis. In other words, this analysis tested whether the intended CPC processes were significantly related to outcomes. If there were not a strong relationship between outcomes and processes, it would be inconsistent with the inference that the processes caused the outcomes. In this demonstration, only one outcome was used (i.e., Prereading Total) but classroom processes were analyzed both individually and as a whole.

The right side of Figure 3 illustrates that the regression of Prereading Achievement on the total of the process measures was statistically significant, indicating that overall the process measures were related to the outcomes. However, the right side of Figure 4 illustrates that when the individual scales were entered in a hierarchical regression analysis (i.e., Child-Centeredness first, Evaluation second, Enriched Environment third, and Parent Involvement fourth), none was found to have a significant relationship with Prereading Achievement. This was interpreted to mean that when a total program including all four of these elements is implemented, it does have the desired effect of improving achievement. In other words, in this study, the individual elements did not seem to contribute significantly to the prediction of Prereading Achievement, but the combination of them did.

Discussion

In this case, the implementation of the CPC program was supported insofar as it was found that the CPC program was strongly related to two of the intended processes and that these processes, taken as a whole, were strongly related to increased achievement. This finding belies the fact that the CPC program (when employed as the sole independent variable) did not show a statistically significant gain in outcomes.

Suchman (1967) and Shapiro (1982) have distinguished between program failure (i.e., the program is not implemented as intended) and theory failure (i.e., the program is implemented as intended but does not cause the expected results). In this case, the program was found to be implemented as intended for two variables but not for two others, thereby indicating some program success. Additionally, the processes, taken as a whole, were found to be significantly related to outcomes, indicating some theory success. Therefore, it may logically be inferred that an intensification of Enriched Environment and Parent Involvement should cause an increase in Prereading Achievement. It was concluded that the theory of the program was, in general, correct, and that the program was well implemented in terms of Child-Centeredness and Evaluation. If there was a failure in the program, it was not due to failure of the causal model, but to the fact that the differences in implementation of the intended program were not great enough between the experimental and comparison programs to provide statistically significant outcomes in the simple regression analysis. To improve the program, the environment should be further enriched relative to the regular program and parent involvement should be increased.

Finally, it should be recalled that this study was viewed as a

prototype useful for the development of an evaluation design not as a crucial experiment. In fact, the study from which these data came had multiple indications of the efficacy of the CPC Program. A program rarely intends only one outcome, so multiple measures are necessary to ensure that the program addresses the interests of a diverse policy shaping community.

Conclusion

In conclusion, this study supported the strategy employed in the Chicago Child Parent Center Program for remediating student achievement. In addition, the program evaluation design which was described here is thought to be a significant improvement in evaluation technology because:

1. It measures implementation of the program as well as outcomes, thereby providing a basis for causal inferences.
2. It provides a method for specifying the theory of the program and a method for testing that theory.
3. It employs available personnel efficiently.
4. It calls upon the available expertise of staff researchers, teachers, school and community representatives and others.
5. It uses much available data productively (e.g., achievement data).
6. It is minimally disruptive of school and classroom processes.
7. With careful sampling, it can provide valid results without an extensive and costly data collection.
8. While providing data useful for decision-making, it also provides data useful for program development and improvement.

Using program implementation measures in path analysis of causal models provides information about the "inner workings" of the program.

Although the problem of measurement error is a constant barrier to drawing causal inferences, the modeling process described here provides substantial benefits in verifying the implementation of treatment and in testing program implementation and program theory as well as program outcomes.

Reference Notes

1. Forgione, P., Kaplan, B., Orland, M. Evaluation of compensatory education programs: Problems, promising strategies and recent trends. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA, 1979.
2. David, J.L. Local uses of Title I evaluations, Stanford Research Institute, Report EPRC 21, Menlo Park, CA, 1978.

References

- Boruch, R.F. & Gomez, H. Measuring impact: Power theory in social program evaluation. In L. Datta & R. Perloff (Eds.), Improving evaluations. Beverly Hills, CA: Sage, 1979.
- Boruch, R.F., Cordray, D.S. An appraisal of educational program evaluations: Federal, state, and local agencies. U.S. Department of Education, 1980.
- Boruch, R.F., Cordray, D.S., Pion, G., & Leviton, L. A mandated appraisal of evaluation practices: Digest of recommendations to the Congress and the Department of Education. Educational Researcher, 1981, 10, 10-13; 31.
- Campbell, D.T. & Boruch, R.F. Making the case for randomized assignment to treatments by considering the alternatives: Six ways in which quasi-experimental evaluations tend to underestimate effects. In C.A. Bennett and A.A. Lumsdaine (Eds.), Evaluation and experience: Some critical issues in assessing social programs. New York: Academic Press, 1975.
- Charters, W.W. and Jones, J.E. On the risk of appraising non-events in program evaluation. Educational Researcher, 1973, 2(11), 5-7.
- Conrad, K.J. & Eash, M.J. Measuring implementation and multiple outcomes in a child parent center compensatory education program. American Educational Research Journal, in press.
- CTB/McGraw-Hill. Comprehensive Tests of Basic Skills, Level A, Form S, Expanded Edition. Monterey, California: CTB/McGraw-Hill, 1973.
- DeVault, M.L., Harnishfeger, A., & Wiley, D. Tools for improving learning: A method for the study of the teaching-learning process. Chicago: Cemrel, Inc., 1978.

- Gersten, R.M., Carnine, D.W. & Williams, P.B. Measuring implementation of a structured educational model in an urban school district: An observational approach. Educational Evaluation and Policy Analysis, 4(1), 1982, 67-80.
- Leinhardt, G. Modeling and measuring educational treatment in evaluation. Review of Educational Research, 1980, 50(3), 393-420.
- Leithwood, K.A. & Montgomery, D.J. A framework for planned educational change: Application to the assessment of program implementation. Educational Evaluation and Policy Analysis, 4(2), 1982, 341-353.
- research. Educational Researcher, 1979, 8, 4-7.
- Lord, F.M. A paradox in the interpretation of group comparison. Psychology Bulletin, 1967, 68, 304-305.
- Rubin, R., Stuck, G., & Revicki, D. A model for assessing the degree of implementation in field-based educational programs. Educational Evaluation and Policy Analysis, 4(2), 1982, 189-196.
- Shapiro, J. Evaluation as theory testing: An example from Head Start. Educational Evaluation and Policy Analysis, 4(3), 1982, 341-353.
- Suchman, E. Evaluative research. New York: Russell Sage Foundation, 1967.
- Walberg, H.J., and Thomas, S.C. Open education: An operational definition and validation in Great Britain and United States. American Educational Research Journal, 9(2), 1972, 197-208.
- Walberg, H.J. and Thomas, S.C. Defining open education. Journal of Research and Development in Education, 8, 1974, 4-13.
- Westinghouse Learning Corporation/Ohio University. The impact of Head Start: An evaluation of the effects of Head Start on children's cognitive and affective development. Washington, D.C.: Office of Economic Opportunity, 1969.

Table 1

The Four A Priori Scales of the "Classroom Observation Rating Scale"
(Walberg and Thomas, 1974) and Their Item Representation

Sample Items	N of Items	Measures Implementation of
1. Child-centeredness:	21	
Children work individually and in small groups at various activities. Children are expected to do their own work without getting help from other children (R). [*] Teacher bases instruction on each individual child and his/her interaction with materials and equipment.		Reduced class size, individualization, humaneness.
2. Evaluation:	6	
Teacher gives children tests to find out what they know. Teacher views evaluation as information to guide instruction and provisioning for the classroom.		Structured basic skills program and evaluation
3. Enriched Environment:	11	
Materials are readily accessible to children. Books are supplied in diversity and profusion (including reference, children's literature). Manipulative materials are supplied in great diversity and range, with little replication.		Provisioning for abundant materials
4. Parent Involvement:	5	
There are parents in the classroom. The environment includes materials for parents to read or use.		Parent involvement in the classroom

* The "R" means that coding was reversed when this item was scored.

Table 2

Group Means, Standard Deviations, and Mean Item Values of CPC vs. non-CPC Classrooms on the Four A Priori Scales of the CORS

Scale	Means ^a		Standard Deviations		Mean Item Values ^b (4 point scale)	
	CPC	non-CPC	CPC	non-CPC	CPC	non-CPC
Child-centeredness	45.4*	32.0	6.3	4.1	2.2	1.5
Evaluation	18.0**	8.5	5.4	2.0	3.1	1.4
Enriched Environment	29.2	26.0	4.9	5.2	2.7	2.4
Parent Involvement	8.2	7.1	2.7	2.7	1.6	1.4

a CPC n=13, non-CPC n=6

b To obtain comparability of measurement on these scales, the group means were divided by the number of items. This gave an indication of relative magnitude on the four point scale.

* $p < .05$

** $p < .01$

Figure 1. A theoretical causal model of the Child Parent Center Program.

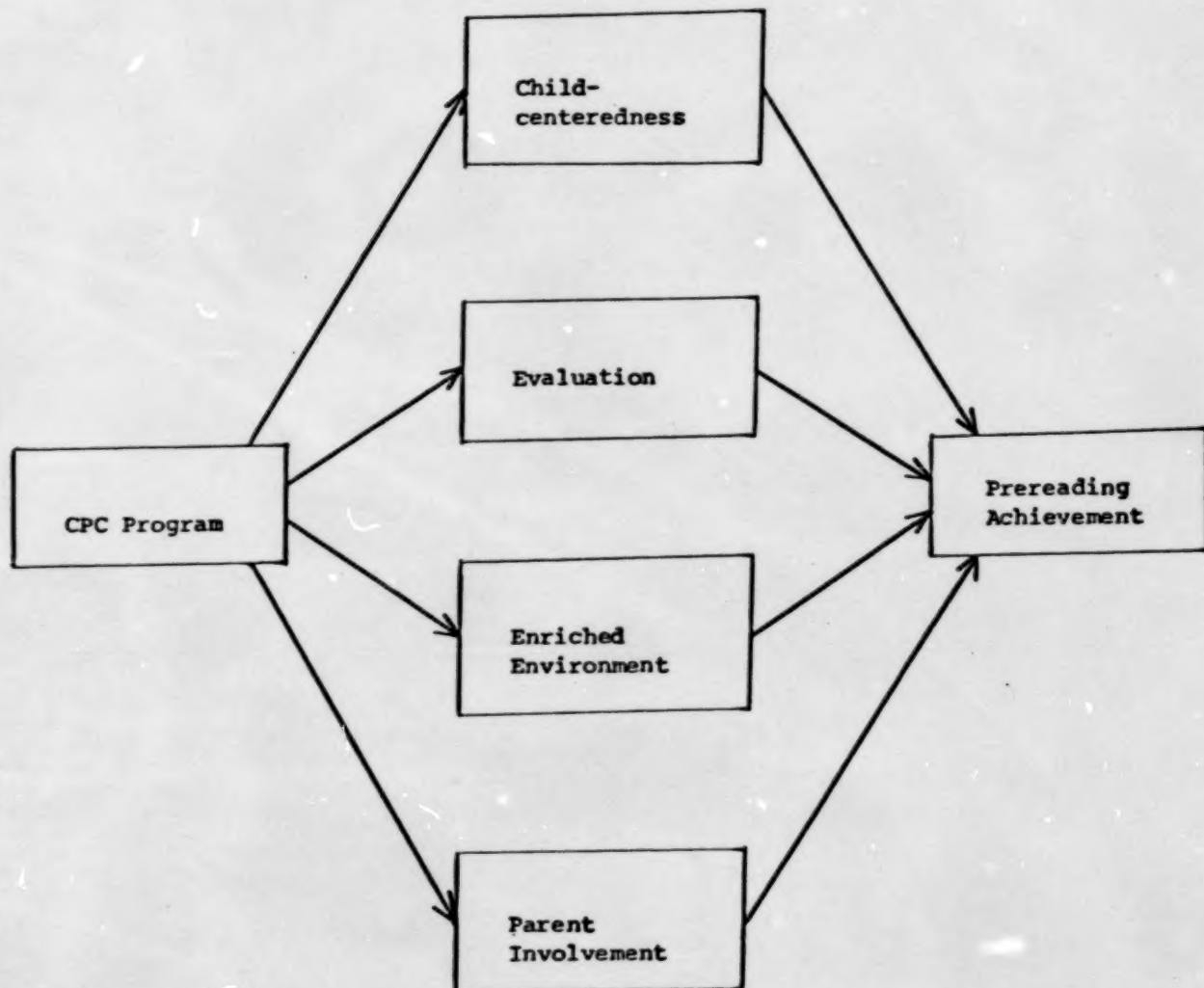


Figure 2. A summative evaluation.

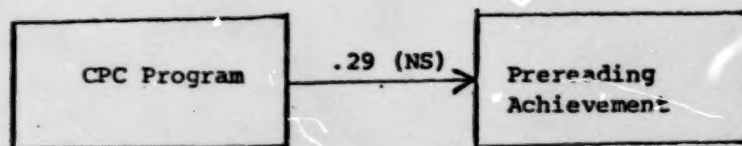
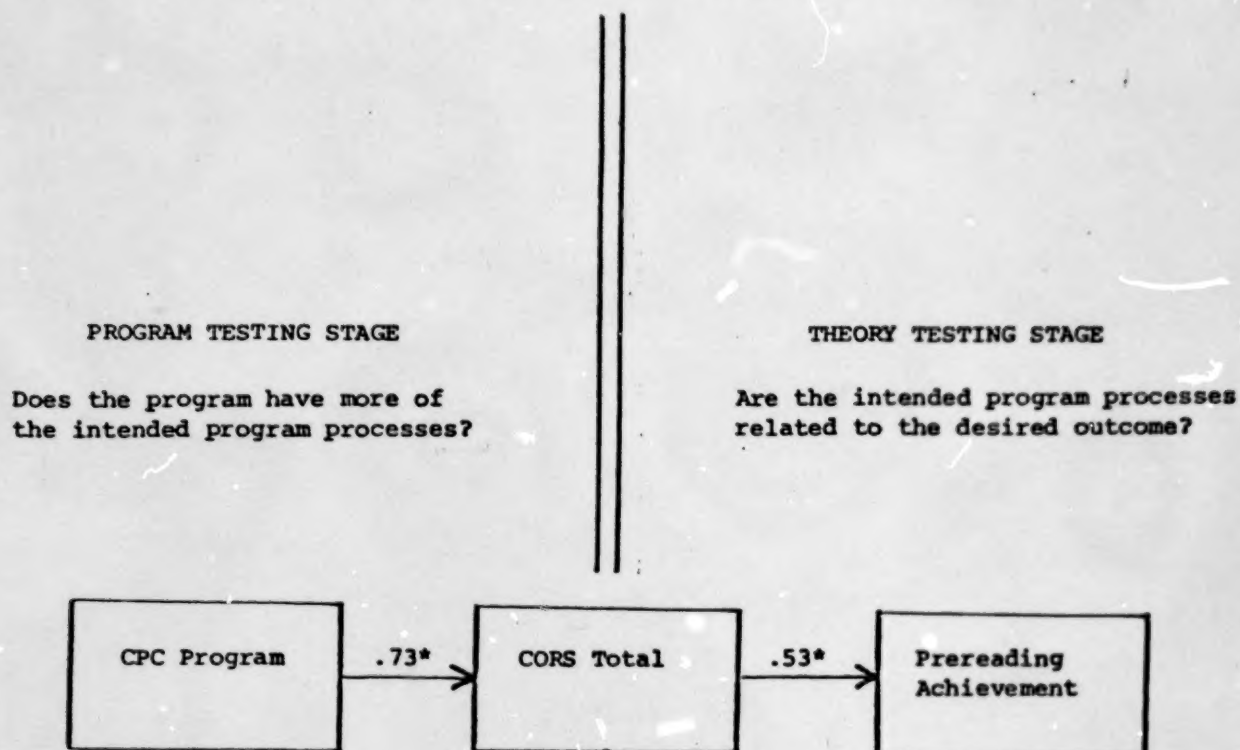
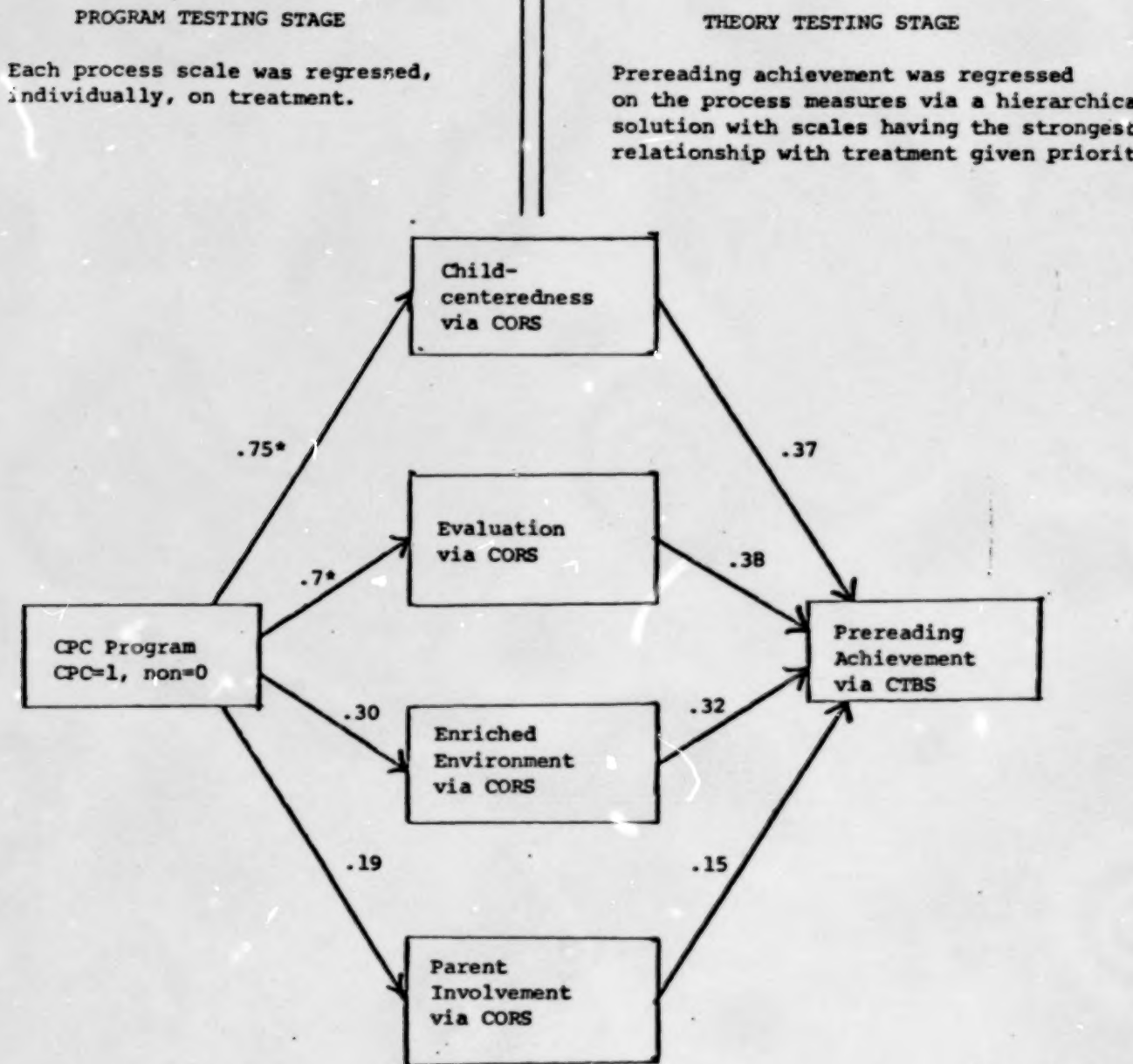


Figure 3. A combination of formative and summative evaluation strategies.



* $p < .05$

Figure 4. An examination of the strength of the relationships of individual processes and treatment and outcomes.



* $p < .05$